STA 104: Take Home Project Part 1
Comparing 2 Drug Groups with Patient Relief from Arthiritis

Aman Singh
amasin@ucdavis.edu
UC Davis
Instructor: Amy T. Kim

2/4/2021

Ⓘ Introduction

Drug treatment has always been a huge topic of research as the FDA seeks many different ways of approving Drugs treatments through trials such as Double Blind Testing, Single Blind Testing, Cross-sectional Studies, Cohort Studies, and many more. In this paper, we will try and analyze whether one Drug treatment provides better relief in the number of hours from Arthritis. This can be very useful as this can give patients a better life with much less pain given the drug. Since the data is not too big in size, I will use an non-parametric approach to find a true conclusion of which drug is more effective.

Ⅱ Summary of Data

This paper uses data from the Drug.1 dataset originating from the STA 104's data sets which has 24 observations with 3 columns titled X,Relief and Drug. We will only be using Relief and Drug in this paper. In the Drug column, there are 2 Groups, one being Drug A, and the other being Drug B.

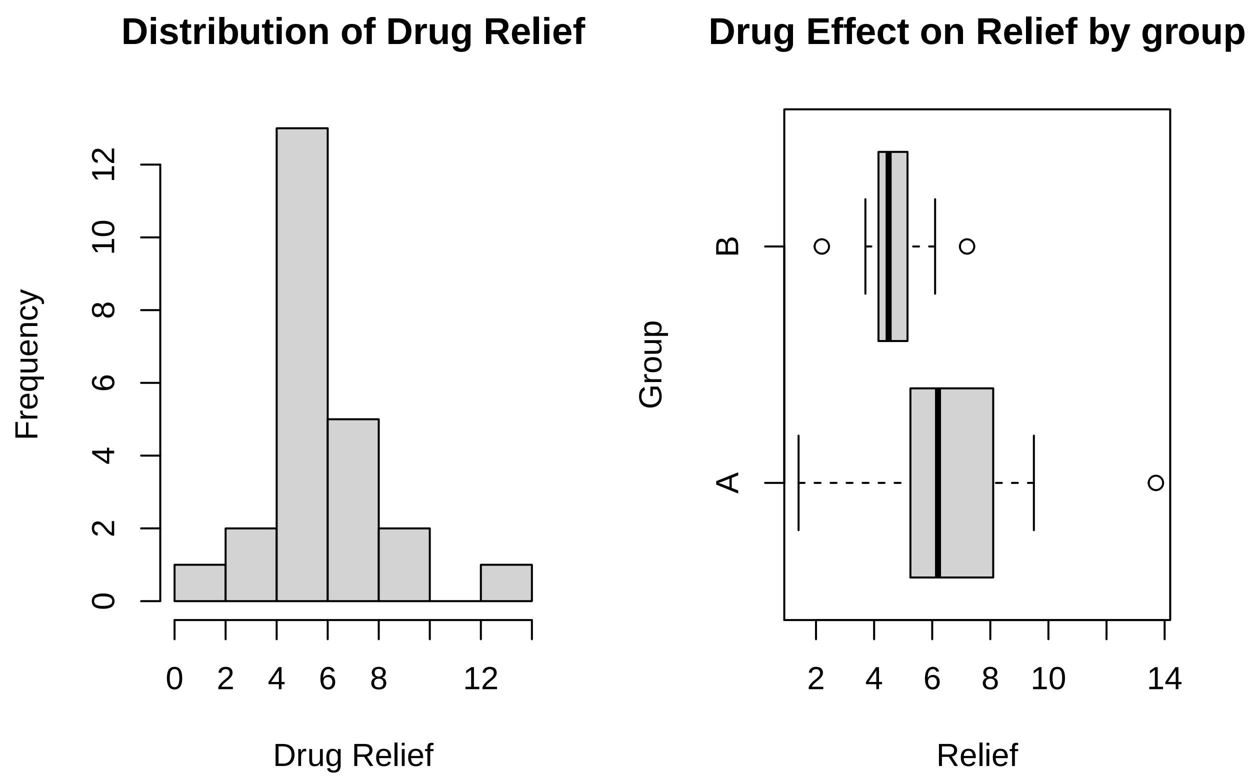

Figure 1: We see an Asymmetrical Distribution in the Histogram. In the Boxplot, we see that there is one outlier in Group A, but other than they are relatively similar in relief, and Group A having a much larger range.

I start by doing some simple analysis on the dataset in order to understand the data a bit more. I found the 5 number summary as well as the mean, standard deviation, and the median on all of the groups and than separately as well. I have included them in the tables below:

Group	Min	25%	Median	75%	Max	Mean	SD	Length
Combined	1.4	4.45	5.20	6.85	13.70	5.725	2.488277	24
Group A	1.4	5.275	6.200	7.700	13.70	6.758333	3.014649	12
Group B	2.200	4.175	4.500	5.125	7.200	4.691667	1.236166	12

Table 1: Summary Statistics:

Given these summary statistics as well as the analysis of the plots, we determine that its best we use an

Nonparametric approach in order to come to an conclusion. The reason is also being the fact that we do not know the distribution of the data hence using an Nonparametric approach.

III Analysis

In the analysis we proceed with using the Mann-Whitney Shift parameter as it would make the most sense in building an confidence interval in order to understand which drug, either A or B, will have the most effect. If the confidence interval ends up being greater than 0, this means that the Group A is more effective than Group B. If the confidence interval ends up being less than 0, this means that Group B is more effective than Group A. If the confidence interval includes 0, we will come to an conclusion that there is no difference between groups A and B.

Percentage	Lower bound	Upper Bound	Result
90%	.6	.3	Group 1 > Group 2
95%	.5	3.3	Group 1 > Group 2
99%	.1	4.4	Group 1 > Group 2

Table 2: Mann-Whitney Shift Parameter Statistics:

We also use the Permutation test for the median to confirm our result from the Shift Parameter statistics. The Permutation test is important because it assigns different values to each value from different outcomes and than calculates it based on the null hypothesis being that the groups have no difference. We also specifically use the median Permutation test given the asymmetrical of the data. We construct the test based on the following hypothesis:

$$H_0 : F_A(x) = F_B(x)$$
$$H_A : F_A(x) \leq F_B(x)$$

We use α being .10,.05 and .01 for testing.

Alpha	P-value	Conclusion
.10	.0064	Reject H_0
.05	.0064	Reject H_0
.01	.0064	Reject H_0

Table 3: Permutation Test of the Median Results:

IV Interpretation

We see that the confidence interval at all three percentages are greater than 0 which concludes that group 1's distribution is greater than group 2's. Group 1 in this case is Drug A being more effective than Drug B for all 3 confidence intervals being 90%, 95% and 99%. We are thus 99% confident that Drug A is more effective than Drug B in providing patients relief from arthritis after taking the respective drug. The shift paramter also shows that the measure of center for Drug A is larger than Drug B.

For the Permutation Test for the Median, we see that we reject the Null hypothesis regardless of α being .10, .05 and .01 thus resulting in the fact that Group A is more effective than Group B.

V Conclusion

After performing the Mann-Whitney Shift parameter as well as the Permutation Test of the medians, we notice that both tests conclude that Drug A is more effective than Drug B. This result can help doctors as well as the FDA in prescribing a drug as well as approving an Drug, respectively. This type of research is incredibly important in having a better control of pain in patients thus leading to a better livelihood which can lead to positive results and an better life for the patient. However, some downsides of this study is that we don't know the price of the Drug as well as the availability of the Drug. This can lead to dramatic differences which can alter the effect of an Drug based on availability.

VI

Code Appendix

```
# cuttingoffcode
library(knitr)
opts_chunk$set(tidy.opts = list(width.cutoff = 70), tidy = TRUE)
Drug.1 <- read.csv("~/Desktop/Statistics/STA 104/Drug-1.csv")
# Part II, summary of data
fivenum(Drug.1$Relief)
tapply(Drug.1$Relief, Drug.1$Drug, summary)
x = mean(Drug.1$Relief)
y = sd(Drug.1$Relief)
z = median(Drug.1$Relief)
x_1 = mean(Drug.1$Relief)
aggregate(Relief ~ Drug, data = Drug.1, mean)
aggregate(Relief ~ Drug, data = Drug.1, sd)
aggregate(Relief ~ Drug, data = Drug.1, median)
aggregate(Relief ~ Drug, data = Drug.1, length)
par(mfrow = c(1, 2))
hist(Drug.1$Relief, xlab = "Drug Relief", main = "Distribution of Drug Relief")
boxplot(Drug.1$Relief ~ as.factor(Drug.1$Drug), xlab = "Relief", ylab = "Group",
    main = "Drug Effect on Relief by group", horizontal = TRUE)
# Part III,Analysis
library(coin)
# 90% Confidence Interval
alpha = 0.1
save.me = wilcox_test(Relief ~ as.factor(Drug), Drug.1, distribution = "exact",
    alternative = "two.sided", conf.int = TRUE, conf.level = 1 - alpha)
confint(save.me)
confint(save.me)$conf.int[1:2]
# 95% Confidence Interval
alpha = 0.05
save.me1 = wilcox_test(Relief ~ as.factor(Drug), Drug.1, distribution = "exact",
    alternative = "two.sided", conf.int = TRUE, conf.level = 1 - alpha)
confint(save.me1)
confint(save.me1)$conf.int[1:2]
# 99% confidence interval
alpha = 0.01
save.me2 = wilcox_test(Relief ~ as.factor(Drug), Drug.1, distribution = "exact",
    alternative = "two.sided", conf.int = TRUE, conf.level = 1 - alpha)
confint(save.me2)
confint(save.me2)$conf.int[1:2]
set.seed(1)
all.perms = sapply(1:10000, function(i) {
    the.numbers = Drug.1$Relief
    the.groups = as.factor(Drug.1$Drug)
    change.groups = sample(the.groups, length(the.groups), replace = FALSE)  # shuffles groups
    group.1.med = median(the.numbers[change.groups == levels(the.groups)[1]])  # finds median for group
    group.2.med = median(the.numbers[change.groups == levels(the.groups)[2]])  # finds median for group
    difference.in.meds = group.1.med - group.2.med  #finds difference in means
    return(difference.in.meds)
})
sample.meds = aggregate(Relief ~ Drug, Drug.1, median)[, 2]  # finds mean per group
```

```
difference = sample.meds[1] - sample.meds[2]  #finds difference in means
p.value.greater = mean(all.perms >= difference)  #calculates p-value for 'greater than' alternative
p.value.greater
```